# CH 3 Hayashi

```
#pmayav2

# clear memory
rm(list=ls())

#import data
grilic <- read.csv("~/Desktop/Econometrics/Assignment 4/3_grilic.csv")
#View(grilic)

col_names <- colnames(grilic)
for( i in 1:length(grilic))
  assign(col_names[i],grilic[[i]])
```

#QUESTION A

```
#Results from table 1
#MEAN
colMeans(grilic[sapply(grilic, is.numeric)])
```

```
##         RNS       RNS80         MRT       MRT80        SMSA      SMSA80
##   0.2691293   0.2928760   0.5145119   0.8984169   0.7044855   0.7124011
##         MED          IQ         KWW        YEAR         AGE       AGE80
##  10.9102902 103.8562005  36.5738786  69.0316623  21.8350923  33.0118734
##           S         S80        EXPR      EXPR80      TENURE    TENURE80
##  13.4050132  13.7071240   1.7354288  11.3942612   1.8311346   7.3627968
##          LW        LW80
##   5.6867388   6.8265554
```

```
#ST. DEV
sapply(grilic, sd, na.rm = TRUE)
```

```
##         RNS       RNS80         MRT       MRT80        SMSA      SMSA80
##  0.4438001  0.4553825  0.5001194  0.3022988  0.4565750  0.4529420
##         MED          IQ         KWW        YEAR         AGE       AGE80
##  2.7411199 13.6186661  7.3022465  2.6317942  2.9817557  3.0855039
##           S         S80        EXPR      EXPR80      TENURE    TENURE80
##  2.2318284  2.2146926  2.1055425  4.2107452  1.6736300  5.0502404
##          LW        LW80
##  0.4289494  0.4099268
```

```
#CORRELATION
cor(IQ,S)
```

```
## [1] 0.5131176
```

#QUESTION B

```
nobs <- length(grilic[[1]])
#dummy variables by year
year_dummies <- matrix(0, ncol=7, nrow=nobs)
year_dummies[which(YEAR==66),1] <- 1
year_dummies[which(YEAR==67),2] <- 1
year_dummies[which(YEAR==68),3] <- 1
year_dummies[which(YEAR==69),4] <- 1
```

```r
year_dummies[which(YEAR==70),5] <- 1
year_dummies[which(YEAR==71),6] <- 1
year_dummies[which(YEAR==73),7] <- 1

excl <- cbind(MED,KWW,MRT,AGE)
h <- cbind(EXPR, TENURE, RNS, SMSA, year_dummies)
```

```r
#line 1 & 2
#using OLS
ols_1 <- lm(LW ~ S + h -1)
ols_2 <- lm(LW ~ S + IQ + h - 1 )
```

```
##
## Call:
## lm(formula = LW ~ S + h - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1703 -0.2193  0.0206  0.2166  1.1439
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## S        0.069673   0.006687  10.419  < 2e-16 ***
## hEXPR    0.029799   0.006524   4.568 5.77e-06 ***
## hTENURE  0.043350   0.007497   5.782 1.08e-08 ***
## hRNS    -0.104092   0.027495  -3.786 0.000165 ***
## hSMSA    0.135267   0.026666   5.073 4.95e-07 ***
## h        4.410852   0.091980  47.955  < 2e-16 ***
## h        4.358365   0.097516  44.694  < 2e-16 ***
## h        4.490244   0.096594  46.486  < 2e-16 ***
## h        4.621975   0.099442  46.479  < 2e-16 ***
## h        4.650202   0.104864  44.345  < 2e-16 ***
## h        4.639710   0.107602  43.119  < 2e-16 ***
## h        4.736954   0.112127  42.246  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3277 on 746 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 1.907e+04 on 12 and 746 DF,  p-value: < 2.2e-16
```

```r
summary(ols_1)
```

```
##
## Call:
## lm(formula = LW ~ S + IQ + h - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1861 -0.2137  0.0204  0.2179  1.1468
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## S        0.061955   0.007279   8.512  < 2e-16 ***
## IQ       0.002712   0.001031   2.630 0.008727 **
```

```
## hEXPR    0.030839   0.006510   4.737 2.60e-06 ***
## hTENURE  0.042163   0.007481   5.636 2.47e-08 ***
## hRNS    -0.096293   0.027547  -3.496 0.000501 ***
## hSMSA    0.132899   0.026576   5.001 7.13e-07 ***
## h        4.235357   0.113349  37.366  < 2e-16 ***
## h        4.181147   0.118223  35.367  < 2e-16 ***
## h        4.315938   0.116838  36.939  < 2e-16 ***
## h        4.442948   0.120193  36.965  < 2e-16 ***
## h        4.463581   0.126281  35.346  < 2e-16 ***
## h        4.458048   0.127515  34.961  < 2e-16 ***
## h        4.558232   0.130741  34.865  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3264 on 745 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 1.774e+04 on 13 and 745 DF,  p-value: < 2.2e-16
```

```
summary(ols_2)
```

```
#line 3
#USING CREATED FUNCTION
#endogenous regressor: IQ
x <- cbind( S, h, excl-1)
z <- cbind(S, IQ, h-1)
y <- LW

source('~/Desktop/Econometrics/Assignment 4/2stageLS_f.R')
twostageLS_f(y, x,  z, flag_print=1)
NA
```

```
##                [,1]
## S       0.0691759100
## IQ      0.0001746559
## EXPR    0.0298660205
## TENURE  0.0432737514
## RNS    -0.1035897013
## SMSA    0.1351148284
##        -0.9224951502
##        -0.9750931618
##        -0.8430265071
##        -0.7115989850
##        -0.6838613459
##        -0.6940342283
##        -0.5966007238
## ********************* 2SLS ***********************
##  Number of Observations: 758
##  Degrees of freedom: 742
##  Mean of the Dependent Variable: 5.686739
##  Variance of the Dependent Variable:: 0.1837548
##  s_sqr: 0.1055649
##  J: 87.65524
##  Sargans: 87.65524
##  P_value: 0
```

```r
library(systemfit)
```

```
## Loading required package: Matrix
## Loading required package: car
## Loading required package: carData
## Loading required package: lmtest
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Please cite the 'systemfit' package as:
## Arne Henningsen and Jeff D. Hamann (2007). systemfit: A Package for Estimating Systems of Simultaneou
##
## If you have questions, suggestions, or comments regarding the 'systemfit' package, please use a forum
## https://r-forge.r-project.org/projects/systemfit/
```

```r
excl <- cbind(MED,KWW,MRT,AGE)
inst_3 <- ~ S + h + excl - 1 #IQ
eqLW <- LW ~ S + IQ + h - 1
#using 2SLS built in function
line_3 <- systemfit(list(eqLW),'2SLS', inst = inst_3 )
summary(line_3)
```

```
##
## systemfit results
## method: 2SLS
##
##          N  DF     SSR  detRCov   OLS-R2 McElroy-R2
## system 758 745 80.0182 0.107407 0.425512   0.425512
##
##       N  DF     SSR      MSE    RMSE       R2   Adj R2
## eq1 758 745 80.0182 0.107407 0.32773 0.425512 0.416258
##
## The covariance matrix of the residuals
##          eq1
## eq1 0.107407
##
## The correlations of the residuals
##     eq1
## eq1   1
##
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: LW ~ S + IQ + h - 1
## Instruments: ~S + h + excl - 1
##
##             Estimate   Std. Error  t value   Pr(>|t|)
## S        0.069175910  0.013048998  5.30124 1.5170e-07 ***
```

```
## IQ       0.000174656  0.003937397  0.04436 0.96463075
## hEXPR    0.029866021  0.006696962  4.45964 9.4753e-06 ***
## hTENURE  0.043273751  0.007693380  5.62480 2.6281e-08 ***
## hRNS    -0.103589701  0.029737132 -3.48351 0.00052378 ***
## hSMSA    0.135114828  0.026888924  5.02493 6.3111e-07 ***
## h        4.399550053  0.270877142 16.24187 < 2.22e-16 ***
## h        4.346952041  0.275143969 15.79883 < 2.22e-16 ***
## h        4.479018696  0.270865964 16.53592 < 2.22e-16 ***
## h        4.610446218  0.278284856 16.56736 < 2.22e-16 ***
## h        4.638183857  0.290522319 15.96498 < 2.22e-16 ***
## h        4.628010975  0.284842041 16.24764 < 2.22e-16 ***
## h        4.725444479  0.282660489 16.71774 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.32773 on 745 degrees of freedom
## Number of observations: 758 Degrees of Freedom: 745
## SSR: 80.01823 MSE: 0.107407 Root MSE: 0.32773
## Multiple R-Squared: 0.425512 Adjusted R-Squared: 0.416258
```

#QUESTION C

```
resid <- resid(line_3)[[1]]
new_x <- cbind( S, h, excl )

Sargans_statistics <-
  function(X,e,n){
  stat <- ( t(e) %*% X %*% solve(t(X) %*% X) %*% t(X) %*% e ) / (t(e) %*% e) * n
  return(stat)}

Sargans_stat <- Sargans_statistics( new_x, resid, nobs )
Sargans_stat
```

```
##          [,1]
## [1,] 87.65524
```

```
#
p_value <- 1-pchisq(Sargans_stat, 2)
p_value
```

```
##      [,1]
## [1,]    0
```

#QUESTION D

```
#Obtaining the 2SLS estimate by running two regressions
#REG1: regressing on IQ to find IQ_hat
i3 <- IQ ~ S + h + excl - 1
IQ_hat <- predict(lm(i3))
#summary(lm(i3))

#REG2: OLS using IQ_hat
d <-  LW ~ S + IQ_hat + h - 1
result_d <- lm(d)
summary(result_d)
```

```
##
```

```
## Call:
## lm(formula = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17039 -0.21960  0.02091  0.21646  1.14359
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## S        0.0691759  0.0130565   5.298 1.54e-07 ***
## IQ_hat   0.0001747  0.0039396   0.044 0.964651
## hEXPR    0.0298660  0.0067008   4.457 9.59e-06 ***
## hTENURE  0.0432738  0.0076978   5.622 2.68e-08 ***
## hRNS    -0.1035897  0.0297542  -3.482 0.000528 ***
## hSMSA    0.1351148  0.0269043   5.022 6.40e-07 ***
## h        4.3995501  0.2710322  16.233  < 2e-16 ***
## h        4.3469520  0.2753014  15.790  < 2e-16 ***
## h        4.4790187  0.2710210  16.526  < 2e-16 ***
## h        4.6104462  0.2784441  16.558  < 2e-16 ***
## h        4.6381839  0.2906886  15.956  < 2e-16 ***
## h        4.6280110  0.2850051  16.238  < 2e-16 ***
## h        4.7254445  0.2828223  16.708  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3279 on 745 degrees of freedom
## Multiple R-squared:  0.9968, Adjusted R-squared:  0.9967
## F-statistic: 1.758e+04 on 13 and 745 DF,  p-value: < 2.2e-16
```

```
#we can see that standard errors are different from B
```

```
#QUESTION E
#line 4
inst_4a <- ~ + h + excl - 1 #IQ
inst_4b <- ~ + h + excl - 1 #S
eqLW_4 <- LW ~ S + IQ + h - 1
eqLW <- LW ~ S + IQ + h - 1

line_4 <- systemfit( list(eqLW, eqLW),"2SLS", inst = list(inst_4a,inst_4b))
#summary(line_4)

e_hat <- resid( line_4 )[[1]]
xx <- cbind( h, excl )

S_stat_f <- Sargans_statistics( xx, e_hat, nobs )
S_stat_f
print(1-pchisq(S_stat_f, 2),digits=20)
```

```
##          [,1]
## [1,] 13.26833
```

#QUESTION E

```
#endogenous regressors: S, IQ
#line 4
x4 <- cbind(h, excl-1) # instruments
```

```
z4 <- cbind(S, IQ, h-1) #dep variables
y4 <- LW

source('~/Desktop/Econometrics/Assignment 4/2stageLS_f.R')
twostageLS_f(y4, x4,  z4, flag_print=1)
NA
```

```
##               [,1]
## S       0.17242531
## IQ     -0.00909883
## EXPR    0.04928949
## TENURE  0.04221709
## RNS    -0.10179345
## SMSA    0.12611095
##        -0.77509264
##        -0.83470975
##        -0.72641306
##        -0.62227500
##        -0.60065661
##        -0.68342667
##        -0.68185287
## ********************* 2SLS ***********************
##  Number of Observations: 758
##  Degrees of freedom: 743
##  Mean of the Dependent Variable: 5.686739
##  Variance of the Dependent Variable:: 0.1837548
##  s_sqr: 0.1418619
##  J: 13.26833
##  Sargans: 13.26833
##  P_value: 0.001314673
```

#QUESTION F

```
#line 5
#GMM
#endogenous regressors: S, IQ
x5 <- cbind( h, excl )
z5 <- cbind(S, IQ, h)
y5 <- LW

source('~/Desktop/Econometrics/Assignment 4/gmm_f.R')
gmm_f(y5, x5, z5, W_hat, flag_print=1)
NA
NA
NA
```

```
##                [,1]
## S       0.175795764
## IQ     -0.009286156
## EXPR    0.050282762
## TENURE  0.042521380
## RNS    -0.104093078
## SMSA    0.124751224
##         4.003924392
##         3.950881214
```

```
##         4.049878999
##         4.159404457
##         4.170911823
##         4.088572858
##         4.103531240
## ********************* GMM function ************************
##  Number of Observations: 758
##  Degrees of freedom : 743
##  Mean of the Dependent Variable: 5.686739
##  Variance of the Dependent Variable: 0.1837548
##  Hansen  J-statistic : 11.60148
##  Chi-squared statistics : 0.003025308
```

```
#2GMM
x_g <- cbind(h, excl,S)
z_g <- cbind(S, IQ, h)
y_g <- LW

source('~/Desktop/Econometrics/Assignment 4/2_gmm_f.R')
efficient2step_gmm_f(y_g, x_g, z_g, flag_print=1)
```

```
##               [,1]
## S       0.17698077
## IQ     -0.01004939
## EXPR    0.04872920
## TENURE  0.04233067
## RNS    -0.10532248
## SMSA    0.12456845
##         4.06913857
##         4.01925099
##         4.11353313
##         4.21465797
##         4.23279170
##         4.16977265
##         4.17547751
##                [,1]
## S       0.076835442
## IQ     -0.001401432
## EXPR    0.031233938
## TENURE  0.048999777
## RNS    -0.100681117
## SMSA    0.133597277
##         4.436784464
##         4.415770982
##         4.525883796
##         4.644032861
##         4.670615269
##         4.671336935
##         4.772811156
## ********************* 2 Step GMM function ************************
##  Number of Observations: 758
##  Degrees of freedom : 742
##  Mean of the Dependent Variable: 5.686739
##  Variance of the Dependent Variable: 0.1837548
##  Hansen  J-statistic1 : 15.99667
```

```
##  Hansen  J-statistic2 : 74.16488
##  C_stat : 58.16822
```

#QUESTION G

```
#endogenous regressors: S, IQ
#line 6
excl6 <- cbind(MRT,AGE)
x6 <- cbind(h, excl6-1) # instruments
z6 <- cbind(S, IQ, h-1) #dep variables
y6 <- LW

twostageLS_f(y6, x6, z6, flag_print=1)

#After excluding MED and KWW; we only use h, MRT, AGE as instruments, which makes the twostageLS_f
#function estimator less precise.
```

```
##             [,1]
## S      -5.292667
## IQ      2.809059
## EXPR    1.894333
## TENURE -1.361397
## RNS     8.947611
## SMSA   -2.902959
##        41.550327
##        39.271898
##        41.945944
##        35.700929
##        26.764504
##        29.986312
##        30.189363
## ********************* 2SLS ***********************
##  Number of Observations: 758
##  Degrees of freedom: 745
##  Mean of the Dependent Variable: 5.686739
##  Variance of the Dependent Variable:: 0.1837548
##  s_sqr: 1062.611
##  J: 6.300423e-15
##  Sargans: 6.300423e-15
##  P_value: 1
```

```
#Yes, the order condition is still satisfied.
```